

TCSS 562: SOFTWARE ENGINEERING FOR CLOUD COMPUTING

Cloud Computing Concepts and Models - II

Wes J. Lloyd
School of Engineering and Technology
University of Washington - Tacoma



1

OFFICE HOURS – FALL 2023

- **Tuesdays:**
 - 2:30 to 3:30 pm - CP 229
- **Fridays**
 - 11:00 am to 12:00 pm – ONLINE via Zoom
- **Or email for appointment**

> *Office Hours set based on Student Demographics survey feedback*

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L7/2
------------------	---	------

2

OBJECTIVES – 10/24

- **Questions from 10/19**
- Tutorials Questions
- Tutorial 5 - to be posted...
- From: Cloud Computing Concepts, Technology & Architecture: Chapter 4: Cloud Computing Concepts and Models:
 - Cloud computing delivery models
 - Cloud deployment models
- AWS Overview and demo
- 2nd hour:
 - Activity 2 - Horizontal Scaling in the Cloud
 - Term Project Planning

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.3
------------------	---	------

3

ONLINE DAILY FEEDBACK SURVEY

- Daily Feedback Quiz in Canvas – Take After Each Class
- Extra Credit for completing

Announcements

Assignments

Discussions

Zoom

Grades

People

Pages

Files

Quizzes

Collaborations

UW Libraries

UW Resources

▼ Upcoming Assignments

- 📎 **Class Activity 1 – Implicit vs. Explicit Parallelism**
Available until Oct 11 at 11:59pm | Due Oct 7 at 7:50pm | -/10 pts
- 📎 **Tutorial 1 - Linux**
Available until Oct 19 at 11:59pm | Due Oct 15 at 11:59pm | -/20 pts

▼ Past Assignments

- 📎 **TCSS 562 - Online Daily Feedback Survey - 10/5**
Available until Dec 18 at 11:59pm | Due Oct 6 at 8:59pm | -/1 pts
- 📎 **TCSS 562 - Online Daily Feedback Survey - 9/30**
Available until Dec 18 at 11:59pm | Due Oct 4 at 8:59pm | -/1 pts

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.4
------------------	---	------

4

TCSS 562 - Online Daily Feedback Survey - 10/5
Started: Oct 7 at 1:13am
Quiz Instructions

Question 1 0.5 pts

On a scale of 1 to 10, please classify your perspective on material covered in today's class:

1	2	3	4	5	6	7	8	9	10
Mostly Review To Me				Equal New and Review					Mostly New to Me

Question 2 0.5 pts

Please rate the pace of today's class:

1	2	3	4	5	6	7	8	9	10
Slow				Just Right					Fast

October 24, 2023 TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma L8.5

5


MATERIAL / PACE

- Please classify your perspective on material covered in today's class (**57** respondents):
 - ew, 10-mostly new
 - **Average - 6.86** (↑ - *previous 6.55*)
- Please rate the pace of today's class:
 - 1-slow, 5-just right, 10-fast
 - **Average - 5.81** (↑ - *previous 5.64*)
- **Response rates:**
 - TCSS 462: 35/44 - 79.5%
 - TCSS 562: 22/25 - 88.0%

October 24, 2023 TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma L8.6

6

FEEDBACK FROM 10/19


Amazon Aurora

- **What Is the difference between Aurora and RDS ?**
- Amazon Aurora is Amazon's fully-managed and hosted relational database service initially released in October 2014.
 - Offered as one of the DB options under Amazon RDS
 - Aurora DBs are either MySQL or PostgreSQL compatible
- Meant to compete with leading edge commercial DBs (i.e. Oracle)
- Storage allocated in 10GB increments up to 128 TB
- Six-way replication of data chunks across three-availability zones provides high availability and fault-tolerance
- Claims of five-fold (5x) performance improvements vs MySQL on identical HW due to custom HW optimizations
 - Rather than improving the DB directly, speed-ups are made possible by through cloud infrastructure optimizations
 - SSD-based virtualized storage layer built for DB workloads
 - Reduces writes to storage system, minimizes lock contention

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.7
------------------	---	------

7

🌐 When poll is active, respond at pollev.com/wesleylloyd641
📧 Text **WESLEYLLOYD641** to **22333** once to join

W The scaling threshold of "when CPU utilization > 80% scale up", is:

An application specific threshold

An application agnostic threshold

Start the presentation to see live content. For screen share software, share the entire screen. Get help at pollev.com/app

8

AWS CLOUD CREDITS UPDATE

- AWS CLOUD CREDITS ARE NOW AVAILABLE FOR TCSS 462/562
- Credits provided on request with expiry of Sept 30, 2024
- Credit codes must be securely exchanged
- Request codes by sending an email with the subject “AWS CREDIT REQUEST” to wllloyd@uw.edu
- Codes can also be obtained in person (or zoom), in the class, during the breaks, after class, during office hours, by appt
 - 41 credit requests fulfilled as of Oct 22 @ 11:59p
- To track credit code distribution, codes not shared via discord
- 52 students have completed AWS Cloud Credits Survey
 - 17 survey responses missing
- **NEXT:** instructor will work to create IAM user accounts
 - One IAM user request (unconfirmed) in queue

October 10, 2023	TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L4.9
------------------	--	------

9

OBJECTIVES – 10/24

- Questions from 10/19
- **Tutorials Questions**
- Tutorial 5 - to be posted...
- From: Cloud Computing Concepts, Technology & Architecture: Chapter 4: Cloud Computing Concepts and Models:
 - Cloud computing delivery models
 - Cloud deployment models
- AWS Overview and demo
- 2nd hour:
 - Activity 2 - Horizontal Scaling in the Cloud
 - Term Project Planning

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.10
------------------	---	-------

10

TUTORIAL 0

- Getting Started with AWS
- http://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_0.pdf
- Create an AWS account
- Create account credentials for working with the CLI
- Install awsconfig package
- Setup awsconfig for working with the AWS CLI

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L7/11
------------------	---	-------

11

TUTORIAL 2 – DUE OCT 21

- **Introduction to Bash Scripting**
- https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_2.pdf
- Review tutorial sections:
- Create a BASH webservice client
 1. What is a BASH script?
 2. Variables
 3. Input
 4. Arithmetic
 5. If Statements
 6. Loops
 7. Functions
 8. User Interface
- Call service to obtain IP address & lat/long of computer
- Call weatherbit.io API to obtain weather forecast for lat/long

October 11, 2022	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L4.12
------------------	---	-------

12

TUTORIAL 3 – DUE OCT 30

- Best Practices for Working with Virtual Machines on Amazon EC2
- http://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_3.pdf
- Creating a spot VM
- Creating an image from a running VM
- Persistent spot request
- Stopping (pausing) VMs
- EBS volume types
- Ephemeral disks (local disks)
- Mounting and formatting a disk
- Disk performance testing with Bonnie++
- Cost Saving Best Practices

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L7/13

13

TUTORIAL 4 – DUE NOV 6

- Introduction to AWS Lambda with the Serverless Application Analytics Framework (SAAF)
- https://faculty.washington.edu/wlloyd/courses/tcss562/tutorials/TCSS462_562_f2023_tutorial_4.pdf (link to be posted)
- Obtaining a Java development environment
- Introduction to Maven build files for Java
- Create and Deploy “hello” Java AWS Lambda Function
 - Creation of API Gateway REST endpoint
- Sequential testing of “hello” AWS Lambda Function
 - API Gateway endpoint
 - AWS CLI Function invocation
- Observing SAAF profiling output
- Parallel testing of “hello” AWS Lambda Function with faas_runner
- Performance analysis using faas_runner reports
- Two function pipeline development task

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.14

14


OBJECTIVES - 10/24

- Questions from 10/19
- Tutorials Questions
- **Tutorial 5 - to be posted...**
- From: Cloud Computing Concepts, Technology & Architecture: Chapter 4: Cloud Computing Concepts and Models:
 - Cloud computing delivery models
 - Cloud deployment models
- AWS Overview and demo
- 2nd hour:
 - Activity 2 - Horizontal Scaling in the Cloud
 - Term Project Planning

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.15
------------------	---	-------

15

CLOUD COMPUTING: CONCEPTS AND MODELS



October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.16
------------------	---	-------

16

OBJECTIVES - 10/24

- Questions from 10/19
- Tutorials Questions
- Tutorial 5 - to be posted...
- **From: Cloud Computing Concepts, Technology & Architecture: Chapter 4: Cloud Computing Concepts and Models:**
 - **Cloud computing delivery models**
 - Cloud deployment models
- AWS Overview and demo
- **2nd hour:**
 - Activity 2 - Horizontal Scaling in the Cloud
 - Term Project Planning

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.17
------------------	---	-------

17

CLOUD COMPUTING DELIVERY MODELS

- **Infrastructure-as-a-Service (IaaS)**
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Serverless Computing:

- Function-as-a-Service (FaaS)
- Container-as-a-Service (CaaS)
- Other Delivery Models

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.18
------------------	---	-------

18

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS) delivery model
- Virtualization is a key-enabling technology of IaaS cloud
- Uses virtual machines to deliver cloud resources to end users

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.19
------------------	---	-------

19

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS) delivery model
- Virtualization is a key-enabling technology of IaaS cloud
- Uses virtual machines to deliver cloud resources to end users

Virtualization is key to sharing powerful servers among users by running many isolated private virtual computers known as virtual machines (VMs)

...VMs are the basis of cloud v1.0

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.20
------------------	---	-------

20


CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS) delivery model
- Virtual Machines
- User
- t

Virtual Machines are the building blocks for “Cloud Service Delivery Models”

They are the “vehicles” used to deliver compute resources to end users...

cloud 1.0







October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.21
------------------	---	-------

21

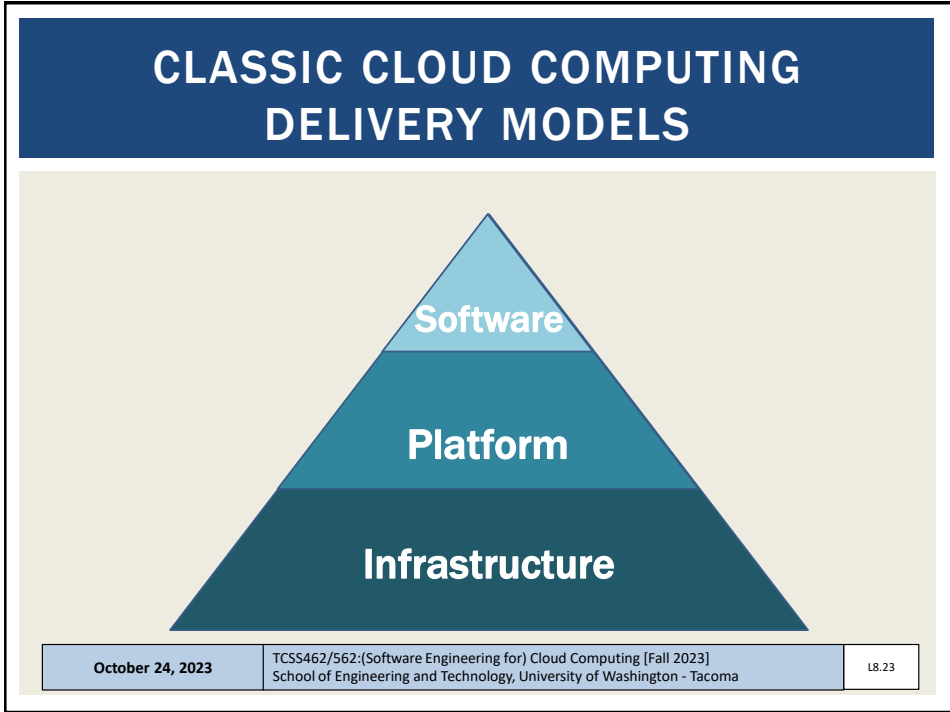
CLOUD COMPUTING DELIVERY MODELS

- What is the appropriate level of **abstraction**?
- How should applications be deployed?
 - IaaS, PaaS, SaaS, DBaaS, FaaS
- How do we ensure Quality-of-Service?
 - Performance, Availability, Responsiveness, Fault Tolerance
- How is **scalability** provided?
- As users, how do we minimize hosting costs?
 - How do we estimate hosting costs?

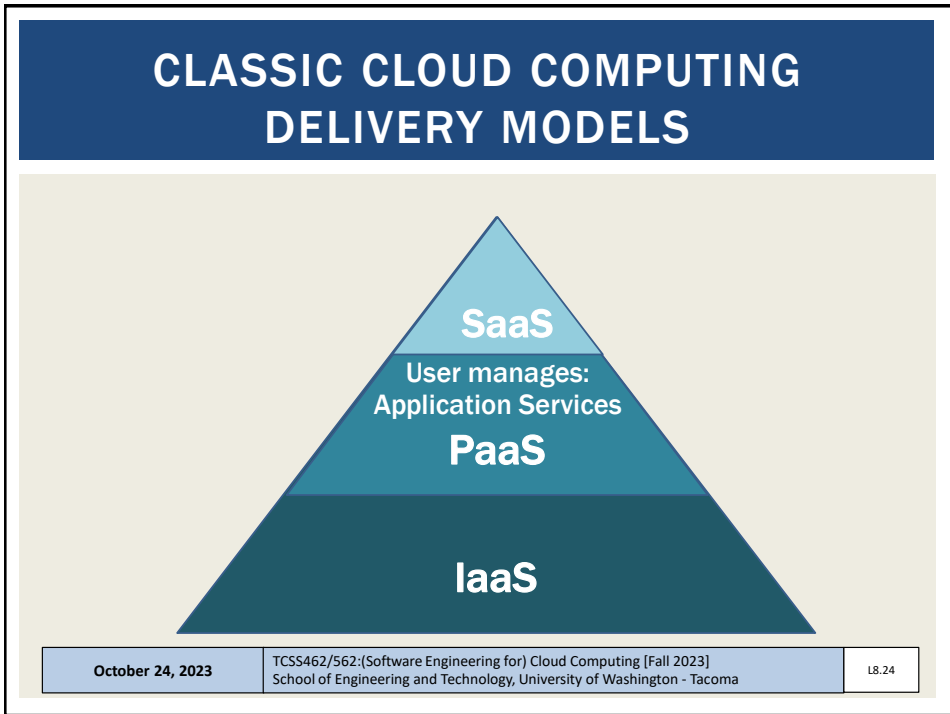


October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.22
------------------	---	-------

22






23



24

EXAMPLE CLOUD SERVICES

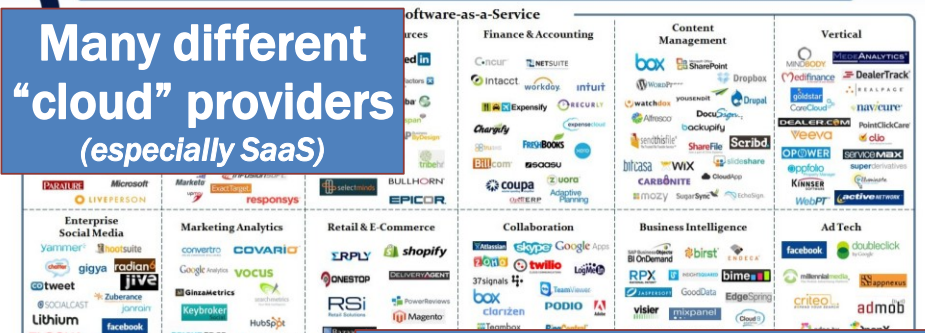
 <h3 style="margin: 0;">SAAS</h3> <p style="margin: 0;">Software as a Service</p> <div style="background-color: #e0e0e0; padding: 5px; margin: 5px;"> Email CRM Collaborative ERP </div> <p style="margin: 0;">CONSUME</p>	 <h3 style="margin: 0;">PAAS</h3> <p style="margin: 0;">Platform as a Service</p> <div style="background-color: #e0e0e0; padding: 5px; margin: 5px;"> Application Development Decision Support Web Streaming </div> <p style="margin: 0;">BUILD ON IT</p>	 <h3 style="margin: 0;">IAAS</h3> <p style="margin: 0;">Infrastructure as a Service</p> <div style="background-color: #e0e0e0; padding: 5px; margin: 5px;"> Caching Legacy Networking Security </div> <div style="background-color: #e0e0e0; padding: 5px; margin: 5px;"> File Technical System Mgmt </div> <p style="margin: 0;">MIGRATE TO IT</p>
--	---	--

October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma
L8.25


25

END USER APPLICATIONS

Many different "cloud" providers (especially SaaS)



Many cloud providers are also cloud consumers



October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma
L8.26

26

INFRASTRUCTURE-AS-A-SERVICE

- Compute resources, on demand, as-a-service
 - Generally raw “IT” resources
 - Hardware, network, containers, operating systems
- Typically provided through virtualization
- Generally, not-preconfigured
- Administrative burden is owned by cloud consumer
- Best when high-level control over environment is needed
- Scaling is generally **not** automatic...
- Resources can be managed in bundles
- **AWS CloudFormation**: Scripts to specify creation of cloud infrastructures using JSON/YAML for app deployment

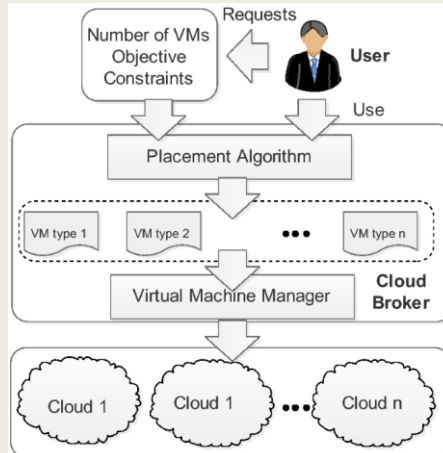
October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.27

27

IAAS: VIRTUAL MACHINE PLACEMENT IN THE CLOUD



Controlled by the cloud provider

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.28

28

COMPONENT COMPOSITION

- Cloud provider maps VMs to physical servers
- User controls mapping of services to VM images
 - Should components be separated (isolated) ?
 - Should components be combined ?
 - M – modeling web service
 - D – relational database
 - F – file server
 - L – logging server

SC1
M D
F L

SC2
M D
F
L

SC3
M D
F L

SC4
M D
F
L

October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
 School of Engineering and Technology, University of Washington - Tacoma
L8.29

29

SC1
M D
F L

SC2
M D
F
L

SC3
M D
F L

SC4
M D
F
L

SC5
M
D
F L

SC6
M
D F
L

SC7
M
D
F
L

SC8
M
D
F L

SC9
M
D L
F

SC10
M F
D L

SC11
M F
D
L

SC12
M L
D F

SC13
M L
D
F

SC14
M D
L
F

SC15
M L
F
D

M: Tomcat ApplicationServer
D: Postgresql DB
F: nginx file server
L: Log server (Codebeamer) ³⁰

30

Bell's Number:

k: number of ways n components can be distributed across containers

n	k
4	15
5	52
6	203
7	877
8	4,140
9	21,147
n	...

M: Tomcat ApplicationServer
 D: Postgresql DB
 F: nginx file server
 L: Log server (Codebeamer) ³¹

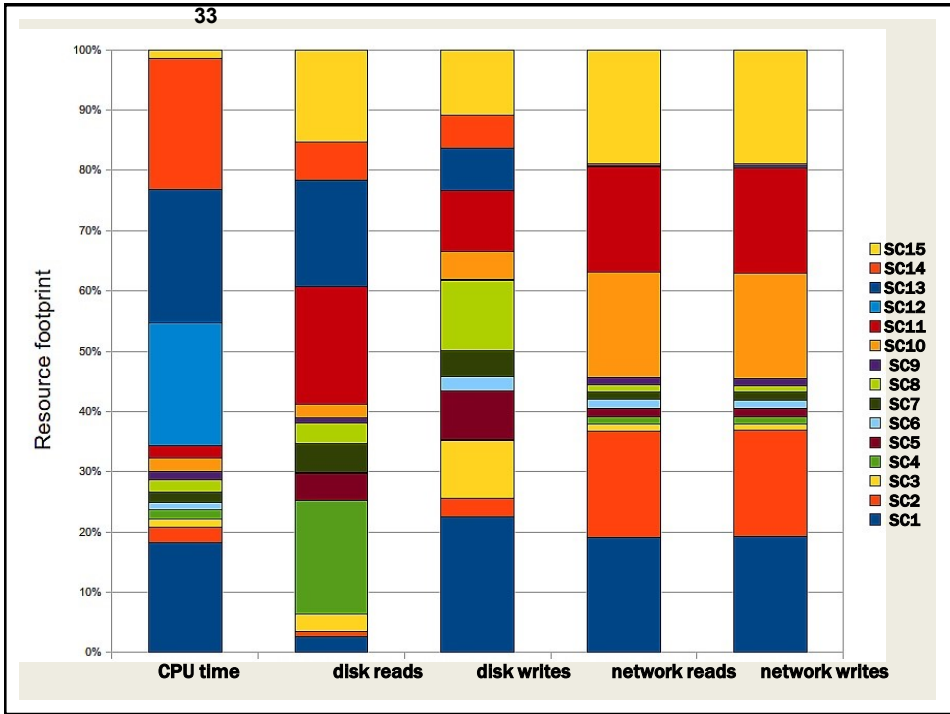
31

Component Composition Example

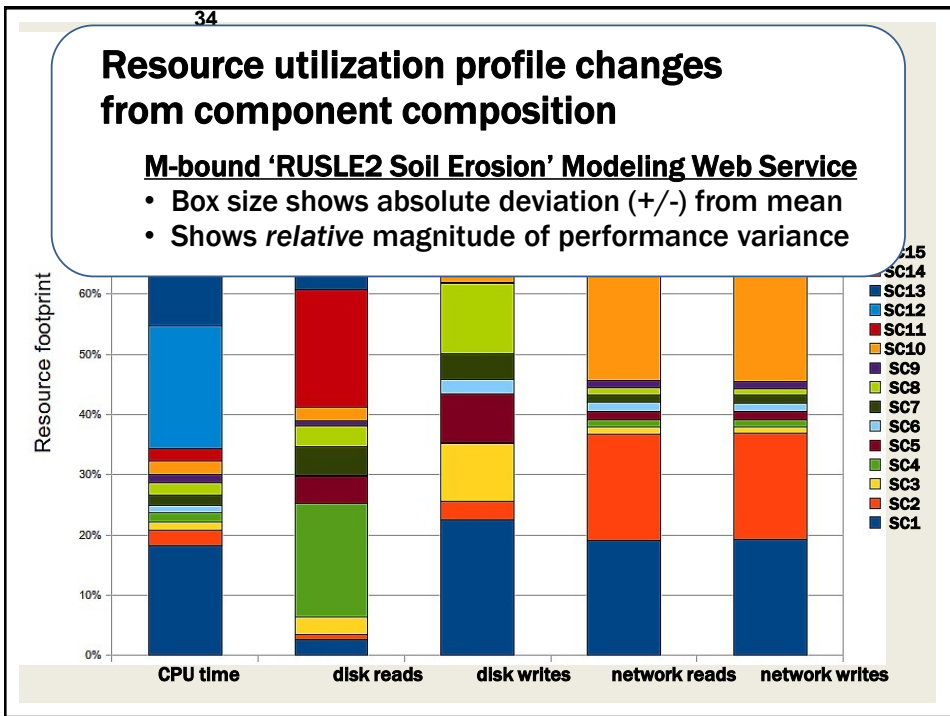
- Application with 4 components has 15 compositions
- One or more component(s) deployed to each VM
- Each VM launched to separate physical machine

M: Tomcat ApplicationServer
 D: Postgresql DB
 F: nginx file server
 L: Log server (Codebeamer) ³²

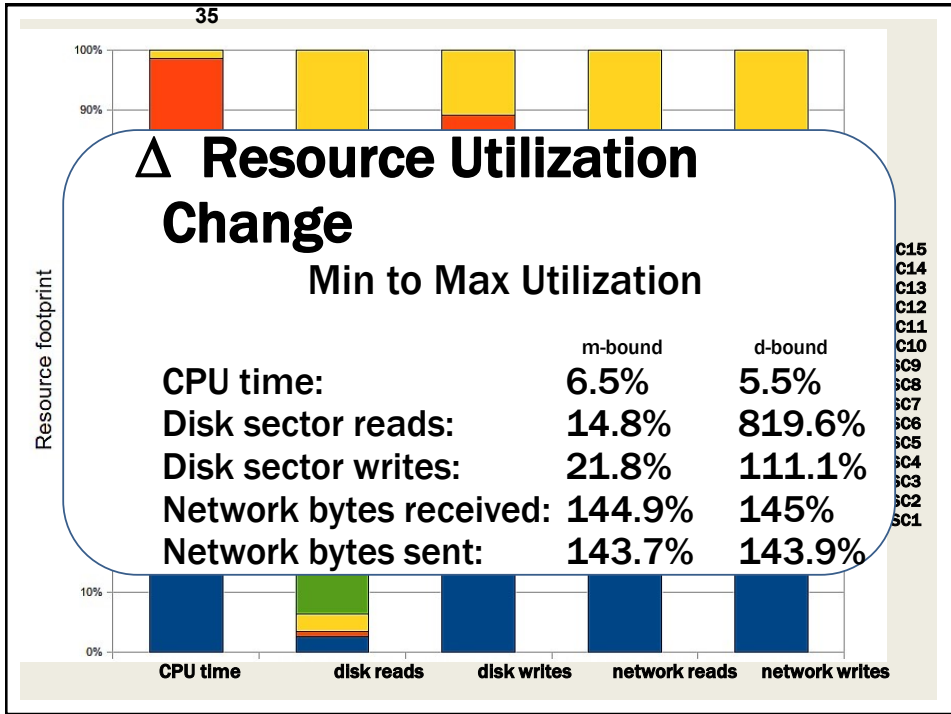
32



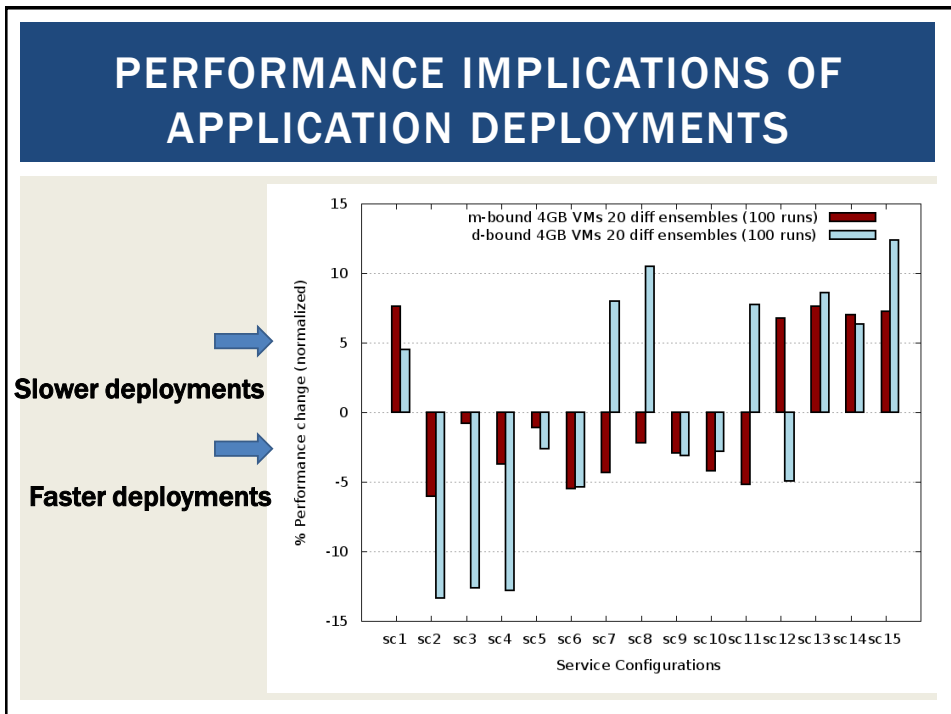
33



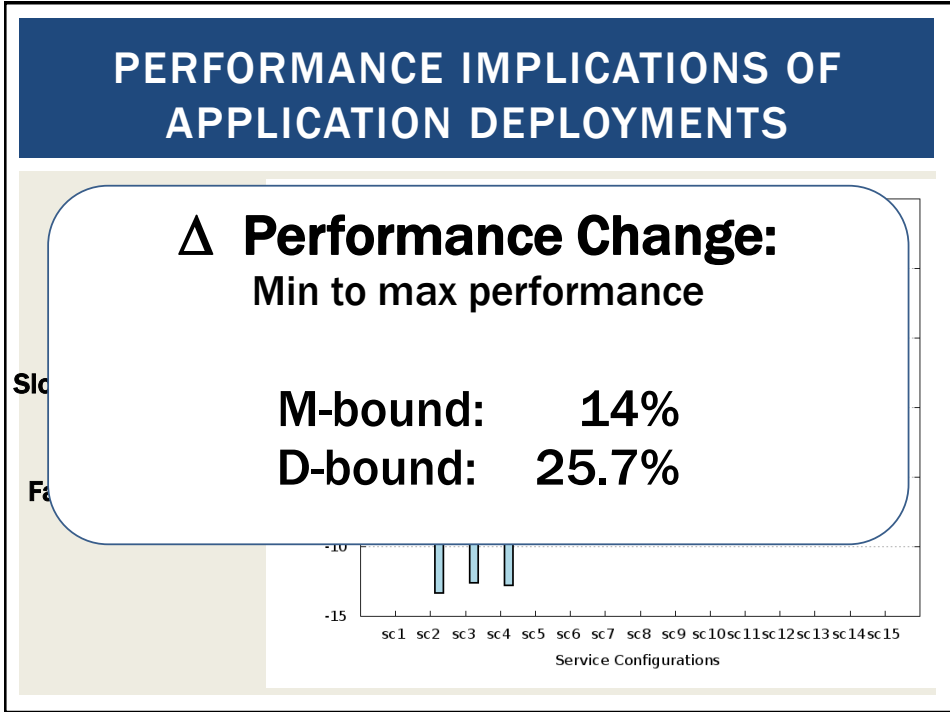
34



35



36



37

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Serverless Computing:

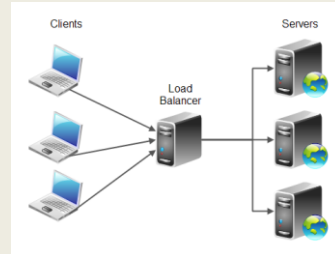
- Function-as-a-Service (FaaS)
- Container-as-a-Service (CaaS)
- Other Delivery Models

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.38
------------------	---	-------

38

PLATFORM-AS-A-SERVICE

- Predefined, ready-to-use, hosting environment
 - Infrastructure is further obscured from end user
 - Scaling and load balancing may be automatically provided and automatic
 - Variable to no ability to influence responsiveness
- Examples:
- Google App Engine
 - Heroku
 - AWS Elastic Beanstalk
 - AWS Lambda (FaaS)



October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.39

39

USES FOR PAAS

- Cloud consumer
 - Wants to extend on-premise environments into the cloud for “web app” hosting
 - Wants to entirely substitute an on-premise hosting environment
 - Cloud consumer wants to become a cloud provider and deploy its own cloud services to external users
- PaaS spares IT administrative burden compared to IaaS

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.40

40

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Serverless Computing:

- Function-as-a-Service (FaaS)
- Container-as-a-Service (CaaS)
- Other Delivery Models

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.41
------------------	---	-------

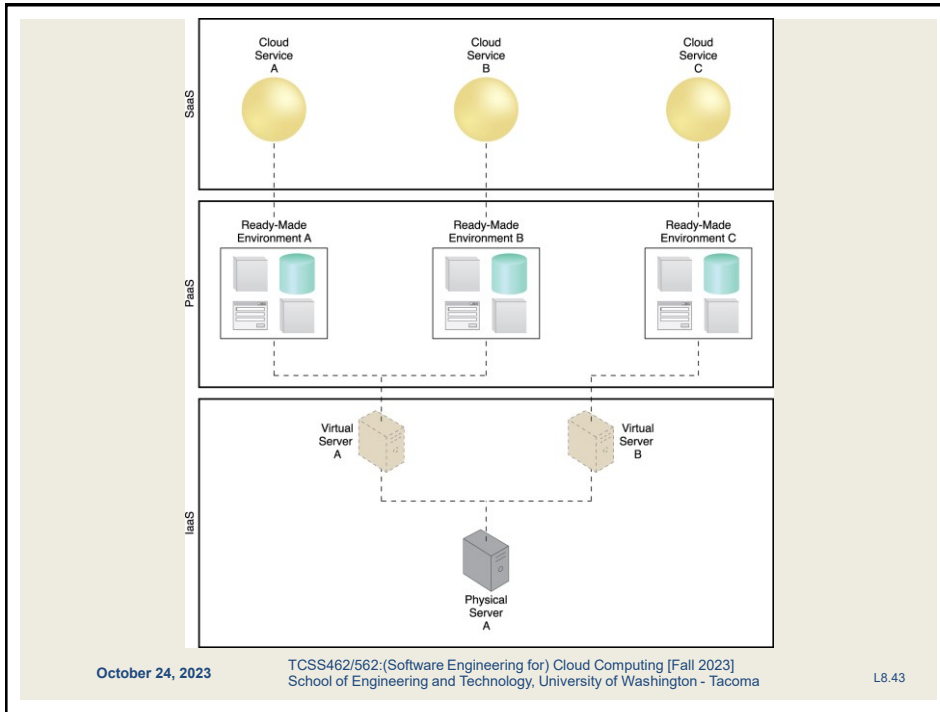
41

SOFTWARE-AS-A-SERVICE

- Software applications as shared cloud service
- Nearly all server infrastructure management is abstracted away from the user
- Software is generally configurable
- SaaS can be a complete GUI/UI based environment
- Or UI-free (database-as-a-service)
- SaaS offerings
 - Google Docs
 - Office 365
 - Cloud9 Integrated Development Environment
 - Salesforce

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.42
------------------	---	-------

42



43

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Serverless Computing:

- Function-as-a-Service (FaaS)
- Container-as-a-Service (CaaS)
- Other Delivery Models

October 24, 2023 TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma L8.44

44

SERVERLESS COMPUTING

Introducing Cloud 2.0



Serverless Computing
Deploy Applications Without Fiddling With Servers

Image from: <https://mobisoftinfotech.com/resources/blog/serverless-computing-deploy-applications-without-fiddling-with-servers/>

45

SERVERLESS COMPUTING

Servers

(AAHHHHHHHHH!!!)

How should my app withstand a server falling?

How can I tell if a server has been compromised?

How can I increase utilization of my servers?

Which OS should my servers run?

How much remaining capacity do my servers have?

How should I implement dynamic configuration changes on my servers?

How will I keep my server OS patched?

When should I decide to scale up my servers?

What size servers are right for my budget?

How can I control access from my servers?

Which packages should be baked into my server images?

How will new code be deployed to my servers?

How many users create too much load for my servers?

What size server is right for my performance?

Which users should have access to my servers?


Should I tune OS settings to optimize my application?

How many servers should I budget for?

When should I decide to scale out my servers?


46

SERVERLESS COMPUTING



What is serverless?

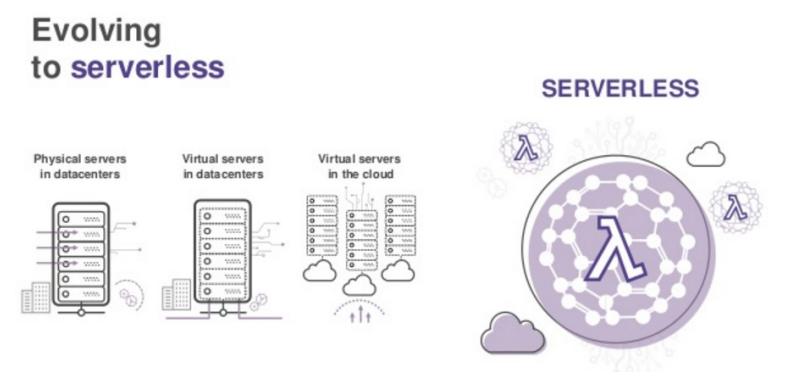
Build and run applications without thinking about servers



October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.47
------------------	---	-------


47

SERVERLESS COMPUTING - 2



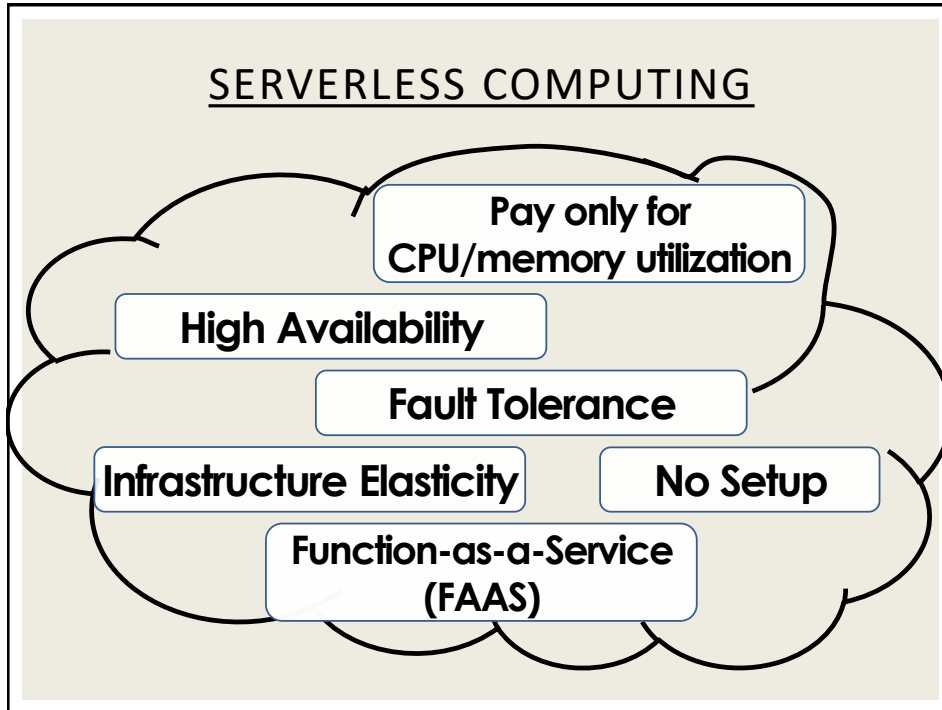
Evolving to serverless

Physical servers in datacenters **Virtual servers in datacenters** **Virtual servers in the cloud** **SERVERLESS**



October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.48
------------------	---	-------

48



49

The slide has a dark blue header with the text "SERVERLESS COMPUTING". Below the header is a light beige background with a central blue rounded rectangle containing the following text:

Why Serverless Computing?
Many features of distributed systems, that are challenging to deliver, are provided automatically
...they are built into the platform

50

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Serverless Computing:

- **Function-as-a-Service (FaaS)**
- Container-as-a-Service (CaaS)
- Other Delivery Models

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.51
------------------	---	-------

51

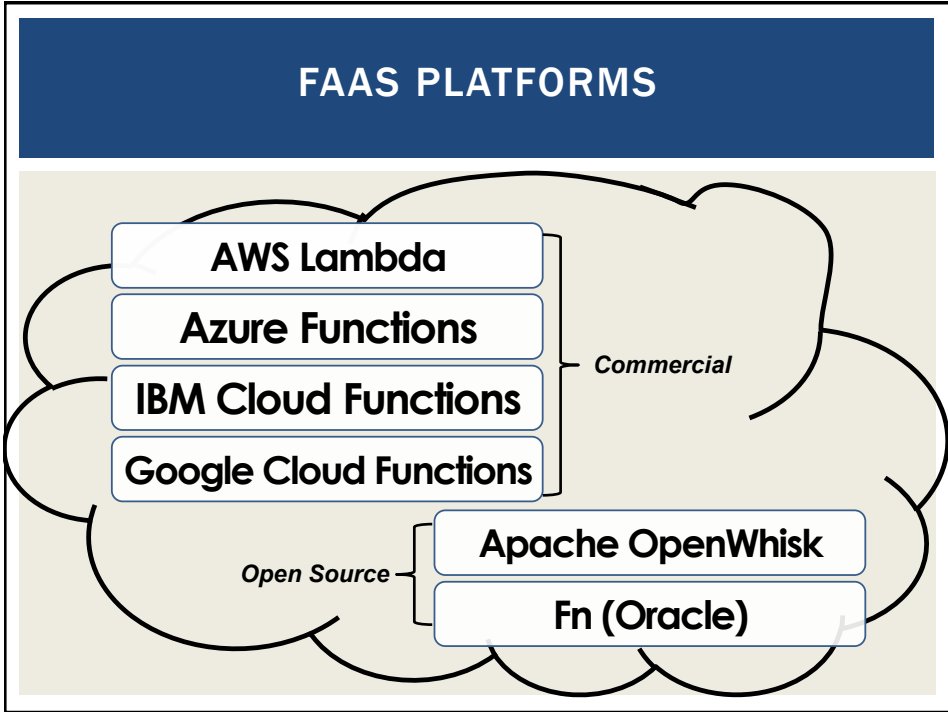
SERVERLESS VS. FAAS

- **Serverless Computing**
- Refers to the avoidance of managing servers
- Can pertain to a number of “as-a-service” cloud offerings
- **Function-as-a-Service (FaaS)**
 - Developers write small code snippets (microservices) which are deployed separately
- Database-as-a-Service (DBaaS)
- Container-as-a-Service (CaaS)
- Others...

- Serverless is a buzzword
- This space is evolving...

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.52
------------------	---	-------

52



53

AWS LAMBDA

Using AWS Lambda

- Bring your own code**
 - Node.js, Java, Python, C#
 - Bring your own libraries (even native ones)
- Simple resource model**
 - Select power rating from 128 MB to 3 GB
 - CPU and network allocated proportionately
- Flexible use**
 - Synchronous or asynchronous
 - Integrated with other AWS services
- Flexible authorization**
 - Securely grant access to resources and VPCs
 - Fine-grained control for invoking your functions

Images credit: aws.amazon.com

54

FAAS PLATFORMS - 2

- New cloud platform for hosting application code
- Every cloud vendor provides their own:
 - AWS Lambda, Azure Functions, Google Cloud Functions, IBM OpenWhisk
- Similar to platform-as-a-service
- Replace opensource web container (e.g. Apache Tomcat) with abstracted vendor-provided **black-box** environment

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.55

55

FAAS PLATFORMS - 3

- Many challenging features of distributed systems are provided automatically
- ***Built into the platform:***
- Highly availability (24/7)
- Scalability
- Fault tolerance

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.56

56

CLOUD NATIVE SOFTWARE ARCHITECTURE

- Every service with a different pricing model

Example: Weather Application

The diagram illustrates a weather application architecture. It starts with S3 (Front-end code for weather app hosted in S3), which is accessed by a laptop (User clicks on link to get local weather information). The laptop sends a REST API call to an API Gateway. The API Gateway triggers Lambda (Lambda is triggered) with a 35°C event. Lambda runs code to retrieve local weather information and returns data back to the user. The data is stored in a database (DYNAMODB).

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.57

57

IAAS BILLING MODELS

- Virtual machines as-a-service at ¢ per hour
- No premium to scale:

$$= \frac{1000 \text{ computers}}{1 \text{ computer}} @ \frac{1 \text{ hour}}{1000 \text{ hours}}$$
- Illusion of infinite scalability to cloud user
- As many computers as you can afford
- Billing models are becoming increasingly granular
 - By the minute, second, 1/10th sec
- Auction-based instances: Spot instances →

The graph shows two data series over time from March 24 to April 03. The y-axis represents cost in dollars, ranging from \$0,000 to \$4,000. One series (red) fluctuates between approximately \$1,500 and \$2,500. The other series (blue) remains relatively flat around \$2,000.

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.58

58

PRICING OBFUSCATION

- **VM pricing:** hourly rental pricing, billed to nearest second is intuitive...
- **FaaS pricing:** non-intuitive pricing policies
- **FREE TIER:**
 - first 1,000,000 function calls/month → FREE
 - first 400,000 GB-sec/month → FREE
- **Afterwards:** *obfuscated pricing (AWS Lambda):*
 - \$0.0000002 per request
 - \$0.000000208 to rent 128MB / 100-ms
 - \$0.00001667 GB /second

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.59
------------------	---	-------

59

WEBSERVICE HOSTING EXAMPLE

- **ON AWS Lambda**
- Each service call: 100% of 2 CPU-cores
100% of 4GB of memory
- Workload: uses 2 continuous threads
- Duration: 1 month (30.41667 days)

- **ON AWS EC2:** Amazon EC2 c5.large 2-vCPU VM x 4GB
- c5.large: 8.5¢/hour, 24 hrs/day x 30.41667 days
- Hosting cost: \$62.05/month

- **How much would hosting this workload cost on AWS Lambda?**

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.60
------------------	---	-------

60

PRICING OBFUSCATION

Assume 1 month = 30.41667 days (365d / 12)

Workload: (4 GB) 10,512,000 GB-sec

Worst-case FaaS scenario = ~2.72x !

AWS EC2:	\$62.05	7
AWS Lambda:	\$168.91	
Break Even:	3,702,459 GB-sec	
@4GB	~10.71 days	3 0

▪ **BREAK-EVEN POINT: \$62.05 - \$0.33 (calls) = \$61.72**
▪ **\$61.72/.00001667 GB-sec = ~3,702,459 GB-sec-mon/4GB/call=**
~925,614 sec or ~10.71 days

Point at which using FaaS costs the same as IaaS

61

FAAS PRICING

- Break-even point is the point where renting VMs or deploying to a serverless platform (e.g. Lambda) is exactly the same.
- Our example is for one month
- Could also consider one day, one hour, one minute
- **What factors influence the break-even point for an application running on AWS Lambda?**

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.62
------------------	---	-------

62

FAAS CHALLENGES

- Vendor architectural lock-in – how to migrate?
- Pricing obfuscation – is it cost effective?
- Memory reservation – how much to reserve?
- Service composition – how to compose software?
- Infrastructure freeze/thaw cycle – how to avoid?
- Performance – what will it be?

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.63
------------------	---	-------

63

VENDOR ARCHITECTURAL LOCK-IN

- Cloud native (FaaS) software architecture requires external services/components

Example: Weather Application

The diagram illustrates a weather application architecture. On the left, a green dollar sign icon is above a bucket icon labeled 'S3'. Below it, text reads 'Front-end code for weather app hosted in S3'. A dashed arrow points from S3 to a laptop icon labeled 'Client'. Below the client, text reads 'User clicks on link to get local weather information'. A dashed arrow points from the client to a red and white striped icon labeled 'API GATEWAY'. Below it, text reads 'App makes REST API call to endpoint'. A dashed arrow points from the API gateway to an orange lambda icon labeled 'Lambda'. Above this arrow, text reads 'Lambda is triggered'. Below the arrow, text reads '35° C'. A dashed arrow points from the lambda icon to a blue database icon labeled 'DYNAMODB'. Below it, text reads 'Lambda runs code to retrieve local weather information and returns data back to user'. A green dollar sign icon is above the lambda icon, and another green dollar sign icon is above the dynamodb icon. A green dollar sign icon is also centered below the lambda icon.

Images credit: aws.amazon.com

- Increased dependencies → increased hosting costs

64


PRICING OBFUSCATION

- **VM pricing:** hourly rental pricing, billed to nearest second is intuitive...
- **FaaS pricing:**
 - AWS Lambda Pricing**
 - FREE TIER:** first 1,000,000 function calls/month → FREE
first 400,000 GB-sec/month → FREE
 - **Afterwards:** \$0.0000002 per request
\$0.000000208 to rent 128MB / 100-ms

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.65
------------------	---	-------

65

MEMORY RESERVATION QUESTION...



- Lambda memory reserved for functions
- UI provides text box formerly “slider bar” to set function’s memory
- Resource capacity (CPU, disk, network) coupled to slider bar:
“every **doubling** of memory, **doubles CPU...**”
- **But how much memory do FaaS functions require?**

▼ Basic settings

Memory (MB) Info
Your function is allocated CPU proportional to the memory configured.

1536 MB

Timeout Info
3 min 0 sec

Description

Performance

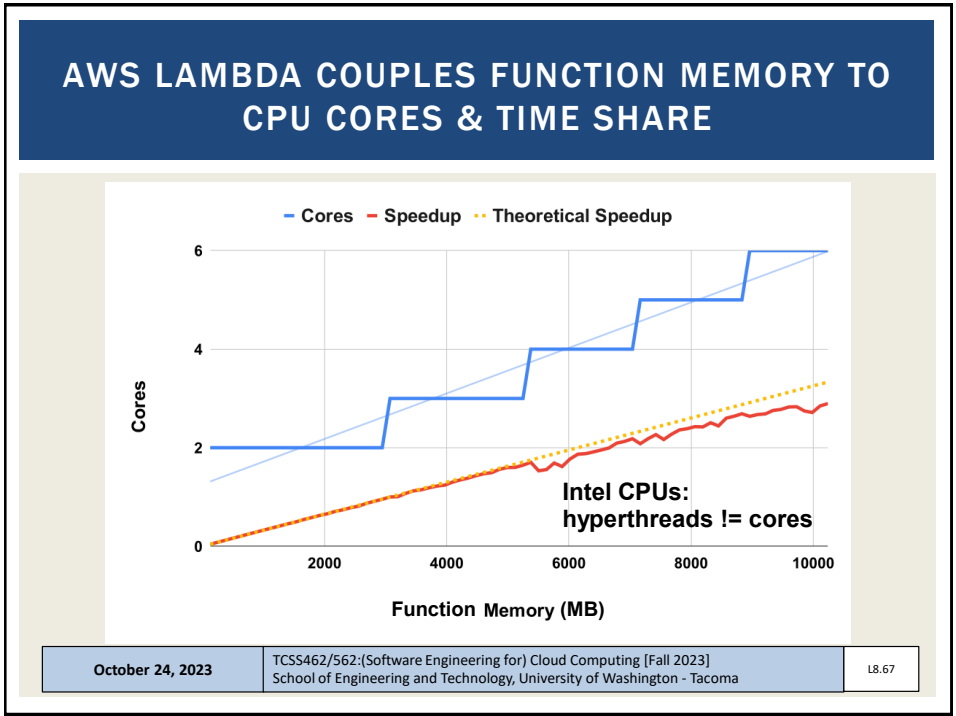
Memory (MB) Info
Your function is allocated CPU proportional to the memory configured.

10240 MB

Set memory to between 128 MB and 10240 MB

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.66
------------------	---	-------

66



67

SERVICE COMPOSITION

- How should application code be composed for deployment to serverless computing platforms?

Monolithic Deployment

Client flow control, 4 functions


Server flow control, 3 functions

- Recommended practice: Decompose into many microservices
- Platform limits: code + libraries ~250MB **Performance**
- How does composition impact the number of function invocations, and memory utilization?

68

INFRASTRUCTURE FREEZE/THAW CYCLE

- **Unused infrastructure is deprecated**
 - *But after how long? (varies by platform)*
- **Infrastructure: microVMs (on AWS Lambda), containers on some platforms**
- **COLD**
 - Code image - built/transferred to physical host & cached
- **WARM**
 - Host has local code cache – create function instance (microVM) on host
- **HOT**
 - Function instance ready to use



Performance


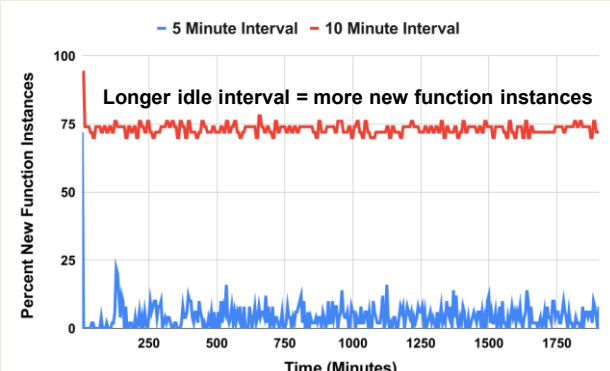


Image from: Denver7 – The Denver Channel News

69

AWS LAMBDA – FREEZE/THAW

- Experiment: 50 concurrent calls, 5 or 10-min calling interval
- Evaluate % cold function instances



— 5 Minute Interval — 10 Minute Interval

Longer idle interval = more new function instances

Percent New Function Instances

Time (Minutes)

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.70
------------------	---	-------

70

FACTORS IMPACTING PERFORMANCE OF FAAS COMPUTING PLATFORMS

- Infrastructure scaling/elasticity
- Resource contention (CPU, network, memory caches)
- Hardware heterogeneity (CPU types, hyperthread, etc)
- Load balancing / provisioning variation
- Infrastructure retention: COLD vs. WARM
 - Infrastructure freeze/thaw cycle
- Function memory reservation size
- Application service composition

October 24, 2023

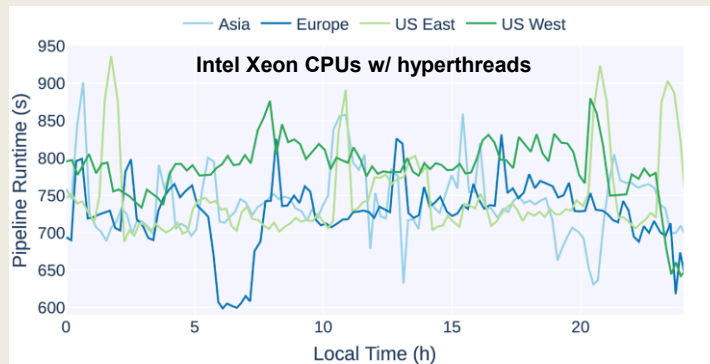
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.71

71

AWS LAMBDA PERFORMANCE VARIATION

- NLP processing pipeline use case
- Performance variance from: diurnal changes in load (e.g. resource contention), Intel hyperthreading

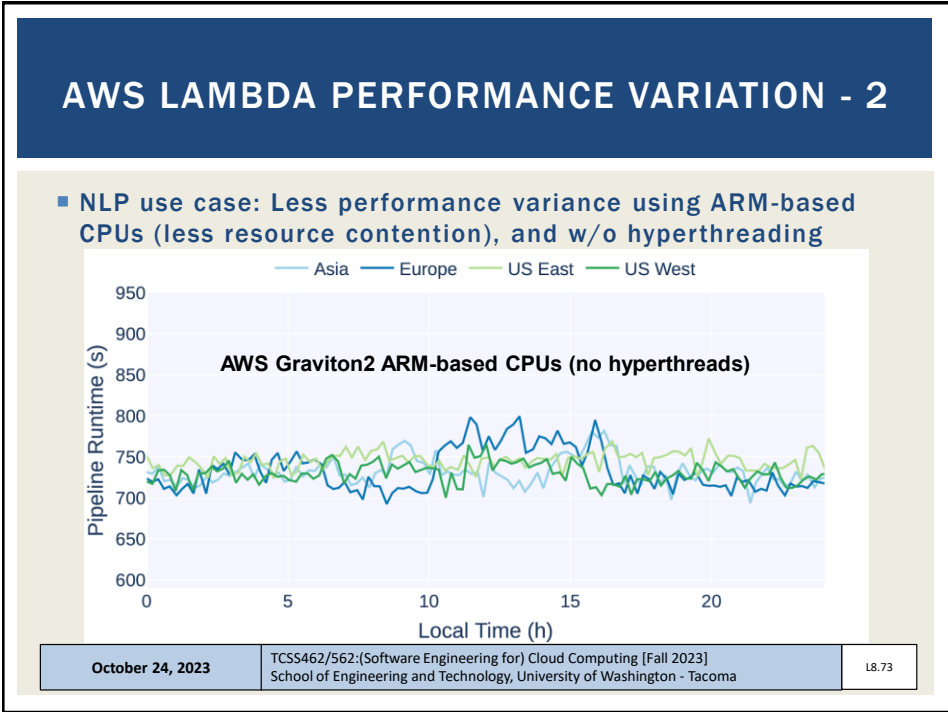


October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.72

72



73



FUNCTION-AS-A-SERVICE

AWS
Lambda
Demo

74

74

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Serverless Computing:

- Function-as-a-Service (FaaS)
- **Container-as-a-Service (CaaS)**
- Other Delivery Models

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.75
------------------	---	-------

75

CONTAINER-AS-A-SERVICE

- Cloud service model for deploying application containers (e.g. Docker containers) to the cloud
- Deploy containers without worrying about managing infrastructure:
 - Servers
 - Or container orchestration platforms
 - Container platform examples: Kubernetes, Docker swarm, Apache Mesos/Marathon, Amazon Elastic Container Service
 - Container platforms support creation of container clusters on the using cloud hosted VMs
- CaaS Examples:
 - AWS Fargate
 - Google Cloud Run
 - Azure Container Instances

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.76
------------------	---	-------

76

CLOUD COMPUTING DELIVERY MODELS

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Serverless Computing:

- Function-as-a-Service (FaaS)
- Container-as-a-Service (CaaS)
- Other Delivery Models

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.77
------------------	---	-------

77

OTHER CLOUD SERVICE MODELS

- IaaS
 - Storage-as-a-Service
- PaaS
 - Integration-as-a-Service
- SaaS
 - Database-as-a-Service
 - Testing-as-a-Service
 - Model-as-a-Service
- ?
 - Security-as-a-Service
 - Integration-as-a-Service

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.78
------------------	---	-------

78

OBJECTIVES - 10/24

- Questions from 10/19
- Tutorials Questions
- Tutorial 5 - to be posted...
- **From: Cloud Computing Concepts, Technology & Architecture: Chapter 4: Cloud Computing Concepts and Models:**
 - Cloud computing delivery models
 - **Cloud deployment models**
- AWS Overview and demo
- **2nd hour:**
 - Activity 2 - Horizontal Scaling in the Cloud
 - Term Project Planning

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.79
------------------	---	-------

79

CLOUD DEPLOYMENT MODELS

- Distinguished by ownership, size, access

- Four common models
 - Public cloud
 - Community cloud
 - Hybrid cloud
 - Private cloud

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.80
------------------	---	-------

80

PUBLIC CLOUDS

organizations

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.81
------------------	---	-------

81

COMMUNITY CLOUD

- Specialized cloud built and shared by a particular community
- Leverage economies of scale within a community
- Research oriented clouds
- Examples:
 - Bionimbus - bioinformatics
 - Chameleon
 - CloudLab

community of organizations

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.82
------------------	---	-------

82

PRIVATE CLOUD

- Compute clusters configured as IaaS cloud
- Open source software
 - Eucalyptus
 - Openstack
 - Apache Cloudstack
 - Nimbus
- Virtualization: XEN, KVM, ...

organization

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.83
------------------	---	-------

83

HYBRID CLOUD

- Extend private cloud typically with public or community cloud resources
- Cloud bursting: Scale beyond one cloud when resource requirements exceed local limitations
- Some resources can remain local for security reasons

organization

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.84
------------------	---	-------

84

OTHER CLOUDS

- **Federated cloud**
 - Simply means to aggregate two or more clouds together
 - Hybrid is typically private-public
 - Federated can be public-public, private-private, etc.
 - Also called inter-cloud
- **Virtual private cloud**
 - Google and Microsoft simply call these virtual networks
 - Ability to interconnect multiple independent subnets of cloud resources together
 - Resources allocated private IPs from individual network subnets can communicate with each other (10.0.1.0/24) and (10.0.2.0/24)
 - Subnets can span multiple availability zones within an AWS region

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.85
------------------	---	-------

85

WE WILL RETURN AT 5:50 PM



86

OBJECTIVES - 10/24

- Questions from 10/19
- Tutorials Questions
- Tutorial 5 - to be posted...
- From: Cloud Computing Concepts, Technology & Architecture: Chapter 4: Cloud Computing Concepts and Models:
 - Cloud computing delivery models
 - Cloud deployment models
- **AWS Overview and demo**
- 2nd hour:
 - Activity 2 - Horizontal Scaling in the Cloud
 - Term Project Planning

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.87
------------------	---	-------

87

AWS OVERVIEW AND DEMO



88

ONLINE CLOUD TUTORIALS

- From the eScience Institute @ UW Seattle:
<https://escience.washington.edu/>
- Online cloud workshops
- Introduction to AWS, Azure, and Google Cloud
- Task: Deploying a Python DJANGO web application
- Self-guided workshop materials available online:
<https://cloudmaven.github.io/documentation/>
- AWS Educate provides access to many online tutorials / learning resources:
<https://aws.amazon.com/education/awseducate/>

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.89
------------------	---	-------

89

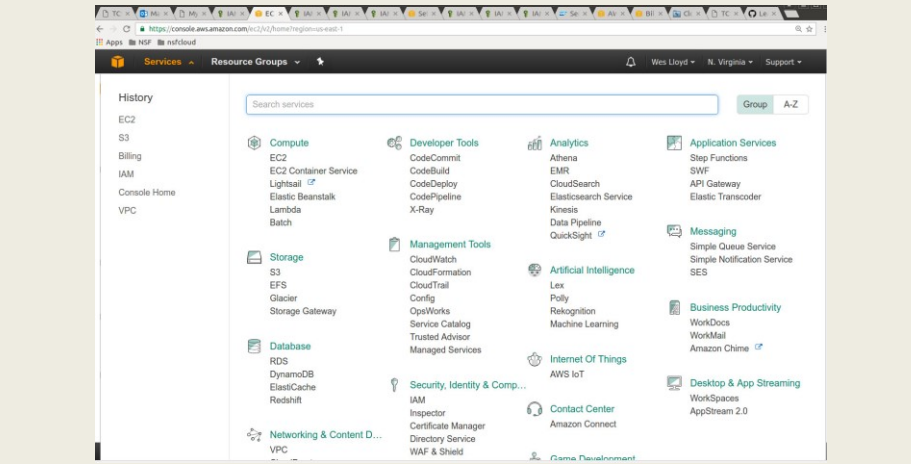
LIST OF TOPICS

- AWS Management Console
- Elastic Compute Cloud (EC2)
- Instance Storage: Virtual Disks on VMs
- Elastic Block Store: Virtual Disks on VMs
- Elastic File System (EFS)
- Amazon Machine Images (AMIs)
- EC2 Paravirtualization
- EC2 Full Virtualization (hvm)
- EC2 Virtualization Evolution
- (VM) Instance Actions
- EC2 Networking
- EC2 Instance Metadata Service
- Simple Storage Service (S3)
- AWS Command Line Interface (CLI)
- Legacy / Service Specific CLIs
- AMI Tools
- Signing Certificates
- Backing up live disks
- Cost Savings Measures

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.90
------------------	---	-------

90

AWS MANAGEMENT CONSOLE



The screenshot shows the AWS Management Console interface. At the top, there's a navigation bar with 'Services' and 'Resource Groups' tabs. Below that, a search bar and a 'Group' dropdown are visible. The main content area is divided into several columns of service tiles, including Compute (EC2, EC2 Container Service, Lightsail, Elastic Beanstalk, Lambda, Batch), Storage (S3, EFS, Glacier, Storage Gateway), Database (RDS, DynamoDB, ElastiCache, Redshift), and many others. A footer bar at the bottom contains the date 'October 24, 2023', the course information 'TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma', and the identifier 'L8.91'.

91

AWS EC2

- **Elastic Compute Cloud**
- **Instance types:** <https://ec2instances.info>
 - **On demand Instance** – full price
 - **Reserved Instance** – contract based where customer guarantees VM rental for a fixed period of time (e.g. 1 year, 3 years, etc.)
Deeper discounts with longer term commitments
 - **Spot Instance** – portion of cloud capacity reserved for low cost instances, when demand exceeds supply instances are randomly terminated with 2 minute warning
 - Users can make diverse VM requests using different types, zones, regions, etc. to minimize instance terminations
 - Developers can design for failure because often only 1 or 2 VMs in a cluster fail at any given time. They then need to be replaced.
 - **Dedicated host** – reserved private HW (server)
 - **Instance families** – General, compute-optimized, memory-optimized, GPU, etc.

October 24, 2023 TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma L8.92

92

AWS EC2 - 2

- Storage types
 - Instance storage - ephemeral storage
 - Temporary disk volumes stored on disks local to the VM
 - Evolution: physical hard disk drives (HDDs)
 - Solid state drives (SSDs)
 - Non-volatile memory express (NVMe) drives (closer to DRAM speed)
 - EBS - Elastic block store
 - Remotely hosted disk volumes
 - EFS - Elastic file system
 - Shared file system based on network file system
 - VMs, Lambdas, Containers mount/interact with shared file system
 - Somewhat expensive

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.93
------------------	---	-------

93

INSTANCE STORAGE

- Also called ephemeral storage
- Persisted using images saved to S3 (simple storage service)
 - ~2.3¢ per GB/month on S3
 - 5GB of free tier storage space on S3
- Requires “burning” an image
- Multi-step process:
 - Create image files
 - Upload chunks to S3
 - Register image
- Launching a VM
 - Requires downloading image components from S3, reassembling them... is potentially slow
- VMs with instance store backed root volumes not pause-able
- Historically root volume limited to 10-GB max- **faster Imaging...**

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.94
------------------	---	-------

94

ELASTIC BLOCK STORE

- EBS provides 1 drive to 1 virtual machine (**1 : 1**) (**not shared**)
- EBS cost model is different than instance storage (uses S3)
 - ~10¢ per GB/month for General Purpose Storage (GP2)
 - ~8¢ per GB/month for General Purpose Storage (GP3)
 - 30GB of free tier storage space
- EBS provides “live” mountable volumes
 - Listed under volumes
 - **Data volumes**: can be mounted/unmounted to any VM, dynamically at any time
 - **Root volumes**: hosts OS files and acts as a boot device for VM
 - In Linux drives are linked to a mount point “directory”
- Snapshots back up EBS volume data to S3
 - Enables replication (required for horizontal scaling)
 - EBS volumes not actively used should be snapshotted, and deleted to save EBS costs...

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.95
------------------	---	-------

95

EBS VOLUME TYPES - 2

- Metric: I/O Operations per Second (IOPS)
- **General Purpose 2 (GP2)**
 - 3 IOPS per GB, min 100 IOPS (<34GB), max of 16,000 IOPS
 - 250MB/sec throughput per volume
- **General Purpose 3 (GP3 – new Dec 2020)**
 - Max 16,000 IOPS, Default 3,000 IOPS
 - GP2 requires creating a 1TB volume to obtain 3,000 IOPS
 - GP3 all volumes start at 3000 IOPS and 125 MB/s throughput
 - 1000 additional IOPS beyond 3000 is \$5/month up to 16000 IOPS
 - 125 MB/s additional throughput is \$5/month up to 1000 MB/s throughput

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.96
------------------	---	-------

96

EBS VOLUME TYPES - 3

- **Provisioned IOPS (IO1)**
 - Legacy, associated with GP2
 - Allows user to create custom disk volumes where they pay for a specified IOPS and throughput
 - 32,000 IOPS, and 500 MB/sec throughput per volume MAX
- **Throughput Optimized HDD (ST1)**
 - Up to 500 MB/sec throughput
 - 4.5 ¢ per GB/month
- **Cold HDD (SC1)**
 - Up to 250 MB/sec throughput
 - 2.5 ¢ per GB/month
- **Magnetic**
 - Up to 90 MB/sec throughput per volume
 - 5 ¢ per GB/month

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.97

97

ELASTIC FILE SYSTEM (EFS)

- EFS provides 1 volume to many client (**1 : n**) **shared storage**
- Network file system (based on NFSv4 protocol)
- Shared file system for EC2, Fargate/ECS, Lambda
- Enables mounting (sharing) the same disk “volume” for R/W access across multiple instances at the same time
- Different performance and limitations vs. EBS/Instance store
- Implementation uses abstracted EC2 instances
- ~ 30 ¢ per GB/month storage – **default burstable throughput**
- **Throughput modes:**
 - Can modify modes only once every 24 hours
- **Burstable Throughput Model:**
 - Baseline – 50Kb/sec per GB
 - Burst – 100MB/sec per GB (for volumes sized 10GB to 1024 GB)
 - Credits - .72 minutes/day per GB

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.98

98

ELASTIC FILE SYSTEM (EFS) - 2

- **Burstable Throughput Rates**
 - Throughput rates: baseline vs burst
 - Credit model for bursting: maximum burst per day

Information subject to revision

File System Size (GiB)	Baseline Aggregate Throughput (MiB/s)	Burst Aggregate Throughput (MiB/s)	Maximum Burst Duration (Min/Day)	% of Time File System Can Burst (Per Day)
10	0.5	100	7.2	0.5%
256	12.5	100	180	12.5%
512	25.0	100	360	25.0%
1024	50.0	100	720	50.0%
1536	75.0	150	720	50.0%
2048	100.0	200	720	50.0%
3072	150.0	300	720	50.0%
4096	200.0	400	720	50.0%

October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma
L8.99

99

ELASTIC FILE SYSTEM (EFS) - 3

- **Throughput Models**
- Provisioned Throughput Model
- For applications with:
 - high performance requirements, but low storage requirements
- Get high levels of performance w/o overprovisioning capacity
- \$6 MB/s-Month (Virginia Region)
 - Default is 50kb/sec for 1 GB, .05 MB/s = 30 ¢ per GB/month
- If file system metered size has higher baseline rate based on size, file system follows default Amazon EFS Bursting Throughput model
 - No charges for Provisioned Throughput below file system's entitlement in Bursting Throughput mode
 - Throughput entitlement = 50kb/sec per GB

Information subject to revision

October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma
L8.100

100

ELASTIC FILE SYSTEM (EFS) - 4

Information subject to revision

	Amazon EFS	Amazon EBS Provisioned IOPS
Per-operation latency	Low, consistent latency.	Lowest, consistent latency.
Throughput scale	10+ GB per second.	Up to 2 GB per second.

Storage Characteristics Comparison, Amazon EFS and Amazon EBS

	Amazon EFS	Amazon EBS Provisioned IOPS
Availability and durability	Data is stored redundantly across multiple AZs.	Data is stored redundantly in a single AZ.
Access	Up to thousands of Amazon EC2 instances, from multiple AZs, can connect concurrently to a file system.	A single Amazon EC2 instance in a single AZ can connect to a file system.
Use cases	Big data and analytics, media processing workflows, content management, web serving, and home directories.	Boot volumes, transactional and NoSQL databases, data warehousing, and ETL.

October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma
L8.101

101

AMAZON MACHINE IMAGES

- AMIs
- Unique for the operating system (root device image)
- Two types
 - Instance store
 - Elastic block store (EBS)
- Deleting requires multiple steps
 - Deregister AMI
 - Delete associated data - (*files in S3*)
- Forgetting both steps leads to costly “orphaned” data
 - No way to instantiate a VM from deregistered AMIs
 - Data still in S3 resulting in charges

October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma
L8.102

102

EC2 VIRTUALIZATION - PARAVIRTUAL

- 1st, 2nd, 3rd, 4th generation → XEN-based
- 5th generation instances → AWS Nitro virtualization

- XEN - two virtualization modes
- XEN Paravirtualization “paravirtual”
 - 10GB Amazon Machine Image – base image size limit
 - Addressed poor performance of old XEN HVM mode
 - I/O performed using special XEN kernel with XEN paravirtual mode optimizations for better performance
 - Requires OS to have an available paravirtual kernel
 - PV VMs: will use common **AKI** files on AWS – **Amazon kernel Image(s)**
 - Look for common identifiers

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.103

103

EC2 VIRTUALIZATION - HVM

- XEN HVM mode
 - Full virtualization – no special OS kernel required
 - Computer entirely simulated
 - MS Windows runs in “hvm” mode
 - Allows work around: 10GB instance store root volume limit
 - Kernel is on the root volume (under /boot)
 - No AKIs (kernel images)
 - Commonly used today (*EBS-backed instances*)

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.104

104

EC2 VIRTUALIZATION - NITRO

- Nitro based on Kernel-based-virtual-machines
 - Stripped down version of Linux KVM hypervisor
 - Uses KVM core kernel module
 - I/O access has a direct path to the device
- **Goal:** provide indistinguishable performance from bare metal

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.105

105

EVOLUTION OF AWS VIRTUALIZATION

- From: <http://www.brendangregg.com/blog/2017-11-29/aws-ec2-virtualization-2017.html>

AWS EC2 Virtualization Types

#	Tech	Type	With	Importance						
				CPU, Memory	Network I/O	Local Storage I/O	Remote Storage I/O	Interrupts, Timers	Motherboard, Boot	
1	VM	Fully Emulated		VS	VS	VS	VS	VS	VS	VS
2	VM	Xen PV 3.0	PV drivers	P	P	P	P	P	VS	VS
3	VM	Xen HVM 3.0	PV drivers	VH	P	P	P	P	VS	VS
4	VM	Xen HVM 4.0.1	PVHVM drivers	VH	P	P	P	P	P	VS
5	VM	Xen AWS 2013	PVHVM + SR-IOV(net)	VH	VH	P	P	P	P	VS
6	VM	Xen AWS 2017	PVHVM + SR-IOV(net, stor.)	VH	VH	VH	P	P	P	VS
7	VM	AWS Nitro 2017		VH	VH	VH	VH	VH	VH	VS
8	HW	AWS Bare Metal 2017		H	H	H	H	H	H	H
		Bare Metal		H	H	H	H	H	H	H

VM: Virtual Machine. HW: Hardware.
 VS: Virt. in software. VH: Virt. in hardware. P: Paravirt. Not all combinations shown.
 SR-IOV(net): igb/ena driver. SR-IOV(storage): nvme driver.

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.106

106

INSTANCE ACTIONS

- Stop
 - Costs of “pausing” an instance
- Terminate
- Reboot

- Image management
- Creating an image
 - EBS (snapshot)
- Bundle image
 - Instance-store

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.107
------------------	---	--------

107

EC2 INSTANCE: NETWORK ACCESS

- Public IP address
- Elastic IPs
 - Costs: in-use FREE, not in-use ~12 ¢/day
 - Not in-use (e.g. “paused” EBS-backed instances)

- Security groups
 - E.g. firewall

- Identity access management (IAM)
 - AWS accounts, groups

- VPC / Subnet / Internet Gateway / Router
- NAT-Gateway

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.108
------------------	---	--------

108

SIMPLE VPC

▪ Recommended when using Amazon EC2

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	igw-id

October 24, 2023
TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma
L8.109

109

VPC SPANNING AVAILABILITY ZONES

Destination	Target
10.0.0.0/16	local

110

INSPECTING INSTANCE INFORMATION

- EC2 VMs run a local metadata service
- Can query instance metadata to self discover cloud configuration attributes
- Find your instance ID:

```
curl http://169.254.169.254/  
curl http://169.254.169.254/latest/  
curl http://169.254.169.254/latest/meta-data/  
curl http://169.254.169.254/latest/meta-data/instance-id  
; echo
```
- `ec2-get-info` command
- Python API that provides easy/formatted access to metadata

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.111

111

SIMPLE STORAGE SERVICE (S3)

- Key-value blob storage
- What is the difference vs. key-value stores (NoSQL DB)?
- Can mount an S3 bucket as a volume in Linux
 - Supports common file-system operations
- Provides eventual consistency
- Can store Lambda function state for life of container.

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.112

112

AWS CLI

- **Launch Ubuntu 16.04 VM**
 - Instances | Launch Instance

- **Install the general AWS CLI**
 - `sudo apt install awscli`

- **Create config file**
[default]

```
aws_access_key_id = <access key id>
aws_secret_access_key = <secret access key>
region = us-east-1
```

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.113
------------------	---	--------

113

AWS CLI - 2

- **Creating access keys: IAM | Users | Security Credentials | Access Keys | Create Access Keys**

October 24, 2023	TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023] School of Engineering and Technology, University of Washington - Tacoma	L8.114
------------------	---	--------

114

AWS CLI - 3

- Export the config file
 - Add to /home/ubuntu/.bashrc

```
export AWS_CONFIG_FILE=$HOME/.aws/config
```
- Try some commands:
 - `aws help`
 - `aws command help`
 - `aws ec2 help`
 - `aws ec2 describes-instances --output text`
 - `aws ec2 describe-instances --output json`
 - `aws s3 ls`
 - `aws s3 ls vmscaleruw`

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.115

115

LEGACY / SERVICE SPECIFIC CLI(S)

- `sudo apt install ec2-api-tools`
- Provides more concise output
- Additional functionality
- Define variables in .bashrc or another sourced script:
 - `export AWS_ACCESS_KEY={your access key}`
 - `export AWS_SECRET_KEY={your secret key}`
- `ec2-describe-instances`
- `ec2-run-instances`
- `ec2-request-spot-instances`
- EC2 management from Java:
 - <http://docs.aws.amazon.com/AWSJavaSDK/latest/javadoc/index.html>
- Some AWS services have separate CLI installable by package

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.116

116

AMI TOOLS

- Amazon Machine Images tools
- For working with disk volumes
- Can create live copies of any disk volume
 - Your local laptop, ec2 root volume (EBS), ec2 ephemeral disk
- Installation:
<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ami-tools-commands.html>
- AMI tools reference:
<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ami-tools-commands.html>
- Some functions may require private key & certificate files

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]
School of Engineering and Technology, University of Washington - Tacoma

L8.117

117

PRIVATE KEY AND CERTIFICATE FILE

- Install openssl package on VM
- ```
generate private key file
$openssl genrsa 2048 > mykey.pk

generate signing certificate file
$openssl req -new -x509 -nodes -sha256 -days 36500 -key
mykey.pk -outform PEM -out signing.cert
```
- Add signing.cert to IAM | Users | Security Credentials |  
- - new signing certificate - -
  - From: [http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/setup-ami-tools.html?icmpid=docs\\_iam\\_console#ami-tools-create-certificate](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/setup-ami-tools.html?icmpid=docs_iam_console#ami-tools-create-certificate)

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L8.118

118

## PRIVATE KEY, CERTIFICATE FILE

- These files, combined with your `AWS_ACCESS_KEY` and `AWS_SECRET_KEY` and `AWS_ACCOUNT_ID` enable you to publish new images from the CLI
- Objective:
  1. Configure VM with software stack
  2. Burn new image for VM replication (**horizontal scaling**)
- An alternative to bundling volumes and storing in S3 is to use a containerization tool such as Docker. . .
- Create image script . . .

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L8.119

119

## SCRIPT: CREATE A NEW INSTANCE STORE IMAGE FROM LIVE DISK VOLUME

```
image=$1
echo "Burn image $image"
echo "$image" > image.id
mkdir /mnt/tmp
AWS_KEY_DIR=/home/ubuntu/.aws
export EC2_URL=http://ec2.amazonaws.com
export S3_URL=https://s3.amazonaws.com
export EC2_PRIVATE_KEY=${AWS_KEY_DIR}/mykey.pk
export EC2_CERT=${AWS_KEY_DIR}/signing.cert
export AWS_USER_ID={your account id}
export AWS_ACCESS_KEY={your aws access key}
export AWS_SECRET_KEY={your aws secret key}
ec2-bundle-vol -s 5000 -u ${AWS_USER_ID} -c ${EC2_CERT} -k ${EC2_PRIVATE_KEY}
--ec2cert /etc/ec2/amiutils/cert-ec2.pem --no-inherit -r x86_64 -p $image -i
/etc/ec2/amiutils/cert-ec2.pem
cd /tmp
ec2-upload-bundle -b tc5562 -m $image.manifest.xml -a ${AWS_ACCESS_KEY} -s
${AWS_SECRET_KEY} --url http://s3.amazonaws.com --location US
ec2-register tc5562/$image.manifest.xml --region us-east-1 --kernel aki-
88aa75e1
```

October 24, 2023

TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L8.120

120

## COST SAVINGS MEASURES

- **From Tutorial 3:**
- **#1: ALWAYS USE SPOT INSTANCES FOR COURSE/RESEARCH RELATED PROJECTS**
- **#2: NEVER LEAVE AN EBS VOLUME IN YOUR ACCOUNT THAT IS NOT ATTACHED TO A RUNNING VM**
- **#3: BE CAREFUL USING PERSISTENT REQUESTS FOR SPOT INSTANCES**
- **#4: TO SAVE/PERSIST DATA, USE EBS SNAPSHOTS AND THEN**
- **#5: DELETE EBS VOLUMES FOR TERMINATED EC2 INSTANCES.**
- **#6: UNUSED SNAPSHOTS AND UNUSED EBS VOLUMES SHOULD BE PROMPTLY DELETED !!**
- **#7: USE PERSISTENT SPOT REQUESTS AND THE “STOP” FEATURE TO PAUSE VMS DURING SHORT BREAKS**

|                  |                                                                                                                                               |        |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|
| October 24, 2023 | TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]<br>School of Engineering and Technology, University of Washington - Tacoma | L8.121 |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|

121

## OBJECTIVES – 10/24

- **Questions from 10/19**
- **Tutorials Questions**
- **Tutorial 5 - to be posted...**
- **From: Cloud Computing Concepts, Technology & Architecture:**
  - **Chapter 4: Cloud Computing Concepts and Models:**
    - **Cloud computing delivery models**
    - **Cloud deployment models**
- **AWS Overview and demo**
- **2<sup>nd</sup> hour:**
  - **Activity 2 - Horizontal Scaling in the Cloud**
  - **Term Project Planning**

|                  |                                                                                                                                               |        |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|
| October 24, 2023 | TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]<br>School of Engineering and Technology, University of Washington - Tacoma | L8.122 |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|

122



## OBJECTIVES - 10/24

- Questions from 10/19
- Tutorials Questions
- Tutorial 5 - to be posted...
- **From: Cloud Computing Concepts, Technology & Architecture: Chapter 4: Cloud Computing Concepts and Models:**
  - Cloud computing delivery models
  - Cloud deployment models
- AWS Overview and demo
- 2<sup>nd</sup> hour:
  - Activity 2 - Horizontal Scaling in the Cloud
  - **Term Project Planning**

|                  |                                                                                                                                               |        |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|
| October 24, 2023 | TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]<br>School of Engineering and Technology, University of Washington - Tacoma | L8.123 |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|

123

# TCSS 462/562 TERM PROJECT



|                  |                                                                                                                                               |        |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|
| October 24, 2023 | TCSS462/562:(Software Engineering for) Cloud Computing [Fall 2023]<br>School of Engineering and Technology, University of Washington - Tacoma | L8.124 |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------|

124

**QUESTIONS**

October 24, 2023

TCSS462/562: (Software Engineering for) Cloud Computing [Fall 2023]  
School of Engineering and Technology, University of Washington - Tacoma

L8.125

125